

ISSN: 1579-9794

La traducción automática de expresiones multiverbales en el turismo gastronómico: un estudio de caso

Machine translation of multiword units in the field of culinary tourism: a case study

ISABEL PEÑUELAS GIL
isabel.penuelas@uva.es
Universidad de Valladolid

Fecha de recepción: 19/02/2024

Fecha de aceptación: 08/10/2024

Resumen: Es innegable que la traducción automática se ha convertido en una constante en el día a día y que ha transformado la forma en que los usuarios abordan el proceso de traducción. Este fenómeno ha tenido un impacto significativo en diversas áreas, especialmente en el contexto del turismo debido a su carácter internacional. Cada vez es más común que empresas, especialmente las de pequeño y mediano tamaño, recurran a herramientas de traducción automática para llegar a un público más amplio y plurilingüe. No obstante, a pesar de su popularidad, estas herramientas pueden ofrecer resultados limitados en términos de calidad y adecuación. El presente trabajo se centra en el estudio de las posibilidades y limitaciones que los sistemas de traducción automática presentan al lidiar con expresiones multiverbales dentro del ámbito del turismo gastronómico. Para ello, se confeccionó un corpus monolingüe (ES), que incluye treinta folletos y guías de diferentes regiones españolas, siguiendo el protocolo de compilación propuesto por Seghiri (2017). A partir de este corpus, se extrajeron las expresiones multiverbales objeto de estudio, junto con sus respectivos contextos, y se sometieron a un proceso de traducción automática utilizando cuatro motores (DeepL, Google Translate, Microsoft Translator y Yandex) pertenecientes a los paradigmas más utilizados hoy en día dentro de la traducción automática para fines específicos. Los resultados obtenidos, categorizados siguiendo una modificación del modelo propuesto por Ortiz Boix (2016), permitieron identificar diferencias de rendimiento entre los sistemas más populares y revelaron los obstáculos comunicativos a los que los usuarios podrían enfrentarse al lidiar con fraseología.

Palabras clave: Traducción automática, Expresiones multiverbales, Turismo gastronómico, Estudio de corpus, Fraseología

Abstract: It is undeniable that machine translation has become a tool that is here to stay and that has transformed the way users approach the translation process. This phenomenon has had a significant impact in several areas, but it is especially noticeable in the context of tourism due to its international nature. It is becoming increasingly common for companies, particularly small and medium-sized ones, to use machine translation tools to reach a wider, multilingual audience. However, despite their popularity, these tools can offer very limited results in terms of quality and appropriateness. This work focuses on the study of the possibilities and limitations that machine translation systems present when dealing with multiword units within the field of culinary tourism. For this purpose, a monolingual corpus (ES) was compiled following the compilation protocol proposed by Seghiri (2017). The corpus includes thirty culinary tourism brochures and guides from different Spanish regions and is the origin of all the multi-word units, as well as their respective contexts. These units were then subjected to a machine translation process using four engines (DeepL, Google Translate, Microsoft Translator, and Yandex), which belong to the most widely used paradigms when it comes to machine translation for specific purposes. The results obtained were categorised following a modified version of the human evaluation system proposed by Ortiz Boix (2016), which allowed to identify the performance differences between some of the most popular engines and revealed the communicative obstacles users might face when dealing with phraseology.

Keywords: Machine translation, Multiword units, Culinary tourism, Corpora studies, Phraseology

INTRODUCCIÓN

Hoy en día, la traducción automática se ha convertido en una constante a nivel mundial con la que convivimos en nuestro día a día. La aparición y popularización de los sistemas de traducción automática (STA) en línea gratuitos en las últimas décadas ha facilitado el proceso de comunicación para muchos (Moorkens, 2022, p. 137). Se trata de herramientas de fácil acceso a nivel mundial, gracias a las cuales una gran cantidad de información que antes podía resultar ilegible para una parte de la población –por la escasez de profesionales que trabajan con un par de lenguas determinado, por cuestiones monetarias o, simplemente, por la cantidad de material para traducir–, ahora está a disposición de la gran mayoría.

Sea cual sea el detonante, no cabe duda de que los STA han facilitado la creación de textos de carácter multilingüe y, en su empeño, han afectado la forma en la que los usuarios interactúan con la traducción. Esto se aplica tanto desde un punto de vista profesional como social, por lo que podemos decir que todos los ámbitos se han visto afectados por estas herramientas de

una manera u otra. Por supuesto, este cambio también se ha dado en el turismo, un ámbito que, tradicionalmente, ya estaba estrechamente ligado a la mediación lingüística, ya fuese en forma de traducción o de interpretación, debido a las características intrínsecas de este.

El turismo es uno de los pilares fundamentales de la economía española. Según datos de 2022, España se sitúa como el segundo país que más turistas internacionales recibe al año, solo por detrás de Francia¹ (Organización Mundial del Turismo [OMT], 2023, p. 6), habiendo recibido 71,6 millones de turistas internacionales a lo largo del año según el Ministerio de Industria, Comercio y Turismo (2023). Este crecimiento del público internacional ha llevado a un aumento de la disponibilidad de traducciones de textos de corte turístico (como pueden ser sitios web, folletos, guías, menús, etc.) destinadas a expandir los servicios y productos proporcionados a un público internacional, de habla no hispana, más amplio.

Por supuesto, una parte de las traducciones son realizadas por profesionales especializados en este sector, pero es común ver que, en el caso de pequeñas y medianas empresas con recursos limitados, estas se elaboran utilizando STA en línea como Google Translate o, más recientemente, DeepL. No obstante, esta práctica puede llevar a problemas de comunicación que no siempre se consideran al decantarse por el uso de la traducción automática (TA) y que contrasta, como indica Álvarez Jurado (2020, p. 4), con el incremento del nivel de exigencia por parte del turista con respecto a los textos que consume.

La relevancia indiscutible que tiene el turismo internacional en la economía española hace crucial prestar atención a las prácticas de los diferentes usuarios que emplean los STA en este sector. Comprender sus necesidades específicas, así como identificar los problemas que puede generar el uso indiscriminado y sin supervisión experta de los STA, permitirá evaluar los efectos que estas tienen en la comunicación y proponer mejoras. Sin embargo, a pesar de esta prominencia, se trata de un área de estudio apenas explorada.

En el presente trabajo, nos centramos en analizar el comportamiento y la efectividad que estas herramientas presentan con respecto al uso de la fraseología, más en concreto de las expresiones multiverbales (EMV), restringiendo la investigación al campo del turismo gastronómico. Como veremos a continuación, estos elementos constituyen una parte fundamental de los textos turísticos, pero también un gran obstáculo desde el punto de

¹ Los datos sobre los movimientos de turistas internacionales durante el 2022 fueron compilados en mayo del 2023. No obstante, la OMT considera que la información que aparece en este informe aún está sujeta a cambios futuros.

vista traductológico y podrían suponer un paso atrás en el intento de mejorar la comunicación con un potencial cliente internacional.

1. LA TRADUCCIÓN AUTOMÁTICA EN PERSPECTIVA

Con la proliferación de los STA en el siglo XXI, también encontramos dos corrientes de pensamiento especialmente prominentes. En las etapas iniciales de la popularización y expansión de estos sistemas, muchos profesionales veían la traducción automática como una amenaza. Como explica Oliver (2016, p. 155), lo hacían bajo la premisa de que la calidad de estos llegaría, en un futuro próximo, a un nivel suficiente como para hacer innecesaria la intervención humana. Por el contrario, el público no experto los percibió de forma positiva, ya que facilitaban el acceso a la información y eliminaban barreras comunicativas (Austermühl y Kortenbruck, 2012, p. 153) «easily, fast and at no ostensible cost» (Vieira, 2020, p. 98). Esta percepción se debe, en parte, a la aparente ausencia de intervención humana, una idea errónea que ha sido reforzada por la prensa popular de manera regular (Bowker, 2023, p. 97), lo que ha llevado al público a malinterpretar el funcionamiento y uso de estas herramientas.

Sin embargo, a pesar de los grandes progresos que se han hecho en el campo de la traducción automática, especialmente con la aparición y auge de los sistemas neuronales, ambas corrientes de pensamiento «resultan optimistas en cuanto a los resultados que estas herramientas pueden ofrecer» (Oliver, 2016, p. 155).

Para entender la realidad a la que se enfrentan los sectores que conviven con un uso regular de la TA, particularmente en el caso del sector turístico, es necesario comprender las carencias de ambos planteamientos arriba expuestos. Ante todo, está la noción, ampliamente aceptada por todos los profesionales de la mediación lingüística, de que este ámbito va más allá de la sustitución de palabras en una lengua (lengua origen, LO) por sus equivalentes en otra (lengua meta, LM). Para que el acto comunicativo que representa una traducción se considere apropiado a la situación, un traductor debe poseer un conocimiento que va más allá del idioma y de las características del lenguaje de especialidad. Este ha de prestar atención a los factores culturales, tal y como expresan Pérez Blanco e Izquierdo (2021, p. 148) al indicar que la propiedad del lenguaje, así como el conocimiento de las convenciones culturales –en todos los niveles– y de tipología textual son fundamentales.

En comparación, observamos que los STA no son capaces de cumplir con los procesos necesarios para alcanzar la aceptabilidad en muchos de esos parámetros establecidos. Aunque la aparición e implementación de nuevas tecnologías en los motores de traducción ha traído mejoras

significativas, especialmente debido al uso de corpus de entrenamiento más grandes –gracias al avance en la capacidad de procesamiento de los ordenadores–, estas tecnologías todavía requieren intervención humana para garantizar la calidad de las traducciones. Como resultado, han surgido nuevas especializaciones, como la preedición y posesición, que buscan optimizar el uso de estas herramientas.

La mera aparición de estas nuevas labores es indicador de que la TA por sí misma no supone una amenaza, a diferencia de lo que algunos profesionales creían inicialmente. No obstante, la ausencia de estas dos labores crea una situación en la que, si un usuario no profesional decide utilizar TA, lo más probable es que opte por usar la traducción resultante sin modificaciones (*raw translation*) y no cuente con una ayuda profesional que la revise.

Es en este contexto donde se manifiestan ciertos problemas que podrían pasar inadvertidos para quienes no están familiarizados con la profesión. Los errores más superficiales, como pueden ser la falta de concordancia en la persona o número gramaticales, son fácilmente detectables si uno traduce a su lengua madre o una segunda lengua en la que se tiene un dominio considerable. Sin embargo, cuestiones como fraseología, humor, dobles sentidos y tonos (como son, por ejemplo, el sarcasmo o la ironía), las cuales tienen la capacidad de modificar por completo el significado de un texto, pueden pasar desapercibidas.

En este trabajo nos centraremos en el análisis de determinadas formas de unidad fraseológica a las que nos referiremos de aquí en adelante como expresiones multiverbales² (Corpas Pastor, 2013) y que entenderemos como combinaciones de al menos dos palabras en las cuales el hablante no tiene libertad total a la hora de elegir sus componentes, puesto que ya han sido convencionalizadas por su uso dentro de un contexto lingüístico determinado (Peñuelas Gil, 2024, p. 73).

Su carácter multipalabra y cultural, que además presenta con frecuencia un sentido idiomático, convierte a las EMV en unos elementos lingüísticos complejos de traducir. Resultan particularmente difíciles en el caso de los traductores automáticos que, como se adelantaba, se limitan únicamente al trasvase entre lenguas. Esto implica que los significados no aparentes de las EMV –es decir, cuando la suma del significado de los

² Existe una falta de unidad por parte de los expertos al referirse al objeto de estudio de la fraseología (Penadés Martínez, 2015, p. 15); dependiendo del autor, se habla de «colocaciones», «expresiones fijas», «fraseologismos», «unidades multiverbales», «frasesmas», etc. Existe una tendencia a utilizar estas denominaciones como sinónimos (Mitkov *et al.*, 2018, p. 3), si bien no es una visión compartida por todos los lingüistas.

componentes no equivale al de la expresión— no se reflejarán en el producto de una traducción automática. Del mismo modo, los posibles dobles sentidos o ambigüedades que puedan surgir, así como las implicaciones culturales de estas expresiones, tampoco estarán presentes en la traducción.

Por otro lado, las EMV son expresiones extremadamente recurrentes que representan un porcentaje significativo del vocabulario habitual de los hablantes nativos de una lengua natural (Jackendoff, 1977; Biber *et al.*, 1999; Sag *et al.*, 2002). Dicha recurrencia implica que el uso apropiado de las mismas resulta indispensable para alcanzar una expresión natural en un lenguaje cualquiera.

Estos factores han hecho que las EMV, más allá de constituir uno de los elementos básicos de los textos turísticos en cualquiera de sus representaciones, también se hayan convertido en uno de los puntos de interés en el campo del procesamiento del lenguaje natural, pues suponen una de las mayores trabas en el progreso de las tecnologías de la traducción:

The successful computational treatment of MWUs [multiword units] is essential for Natural Language Processing, including Machine Translation and Translation Technology; the inability to detect MWUs automatically may result in the incorrect (and even unfortunate) automatic translation and may jeopardise the performance of applications. (Mitkov *et al.*, 2018, p. 3)

Dado el crecimiento de la traducción automática de textos turísticos y las reflexiones arriba representadas, no es descabellado afirmar que el presente y futuro de ambos ámbitos están estrechamente ligados al avance en cuestiones del procesamiento de las EMV.

1.1. *Procesamiento de expresiones multiverbales en la traducción automática*

A medida que los sistemas de TA han avanzado, también lo ha hecho la investigación sobre cómo los distintos paradigmas se enfrentan a las EMV y a los resultados obtenidos. Cada paradigma procesa de manera diferente la información lingüística con la que se crea, así como los textos de entrada que se van a someter al proceso de traducción, lo que modifica inevitablemente el resultado obtenido. En la actualidad, entre los STA más comunes en la traducción para fines específicos, encontramos los sistemas estadísticos, neuronales e híbridos, más conocidos como SMT, NMT y HMT, respectivamente, por sus siglas en inglés.

Tanto los SMT como los NMT requieren del uso de corpus para su entrenamiento. Por su parte, y por norma general, los motores híbridos actuales también basan su arquitectura en el uso de corpus, ya que son el

producto de la combinación de resultados de otros paradigmas (Costa-Jussà y Fonollosa, 2015, p. 4). Tradicionalmente, estos motores HMT se han basado de forma parcial en los sistemas SMT y, más recientemente, han aprovechado asimismo la tecnología neuronal. Por lo tanto, los tres paradigmas parten, en la mayoría de los casos, de la tecnología de corpus. No obstante, a pesar de este punto en común, todos ellos procesan los datos de manera diferente.

Los sistemas estadísticos han significado un gran progreso en cuestiones de TA debido a la relación positiva entre la calidad de los resultados y el esfuerzo necesario para crear los sistemas. Los SMT dominaron el mercado hasta la aparición de la tecnología neuronal; no obstante, hoy en día, siguen teniendo su utilidad en ámbitos especializados. Como su propio nombre indica, extraen información estadística de los grandes corpus utilizados en la fase de entrenamiento del motor (Brown *et al.*, 1988, p. 71) y se «sustentan en la probabilidad (alta o baja) de que a una oración en la lengua de origen le corresponda una traducción en la lengua meta» (Sánchez Ramos y Rico Pérez, 2020, p. 15).

Por supuesto, este es el mismo principio que utilizan para la traducción de EMV. Sin embargo, las bases teóricas de este enfoque –primero basado en la alineación de pares de palabras (Brown *et al.*, 1988, 1990) y, posteriormente, basado en la alineación de frases (Zens *et al.*, 2002; Koehn *et al.*, 2003)– no atienden a las necesidades específicas de la traducción de EMV.

El primer caso presenta dos obstáculos: por un lado, la discordancia entre el carácter idiomático subyacente de las EMV y la división palabra por palabra de la alineación que entiende todos los elementos constituyentes de una EMV como elementos individuales; y, por el otro, su incapacidad para procesar correspondencias «many-to-many» (Mitkov *et al.*, 2018, p. 16).

Este segundo obstáculo queda solventado en el modelo basado en la alineación de frases. Sin embargo, estas frases, más conocidas como n-gramas, no siempre equivalen a EMV y pueden resultar en combinaciones con una importancia limitada desde el punto de vista lingüístico (p. ej.: «la jornada»). Se trata de un problema que motores SMT con otros modelos de alineación, específicamente aquellos que incluyen información sintáctica (Yamada y Knight, 2001), mejoran y es la tendencia que se ha seguido en los últimos años. No obstante, como señalan Mitkov *et al.*:

In the state-of-the-art PB-SMT [phrase-based statistical machine translation] systems, the correct translation of MWUs occurs therefore only on a statistical basis if the constituents of MWUs are marked and aligned as parts of consecutive phrases (n-grams) in

the training set and it is not generally treated as a special case where correspondences between source and target may not be so straightforward, i.e. it does not consist of consecutive many-to-many source-target correspondences (Mitkov *et al.*, 2018, p. 17).

Los sistemas neuronales, por su parte, despuntaron a mediados de la década de 2010 y se presentaron no solo como uno de los paradigmas más innovadores, sino como un paradigma capaz de progresar rápidamente (Sánchez Ramos y Rico Pérez, 2020, p. 12). Se basan en redes neuronales artificiales, un modelo computacional que «comprises a large number of highly interconnected processing elements that work in unison to solve specific problems» (Bowker y Buitrago Ciro, 2019, p. 45).

Al tratarse de sistemas relativamente nuevos, la investigación de cómo estos gestionan y traducen las EMV es aún limitada. Pero cuentan con una ventaja potencial en este aspecto: los sistemas neuronales aprenden al traducir. Utilizan una tecnología de aprendizaje automático que hace que la red neuronal sea capaz de identificar patrones complejos en los textos de entrenamiento³ y los utiliza para aprender a traducir nuevos textos de forma automática (Bowker, 2023, p. 96).

Desde el punto de vista de las EMV, esto implica que, con el tiempo, los motores NMT podrían aprender a traducir cualquier expresión. Sin embargo, para llegar a ese punto, es necesario hacer un uso adecuado de estas herramientas para que no vuelvan a aprender los mismos errores, perpetuándolos así. Se trata de una noción de gran importancia y novedosa en el campo que lleva a reflexiones como la que hace Kenny:

A NMT system might indeed produce an idiomatic translation, but this is generally because the data it has learned from contain hundreds or maybe thousands of examples of that very translation. An NMT system [...] does not know it is being idiomatic, or using a cultural equivalent, when it correctly translates [an idiom] (Kenny, 2022, p. 39).

Es decir, aunque estos sistemas han supuesto un avance en el procesamiento de la información, su forma de «comprender» los textos no ha cambiado sustancialmente. Como señala Kenny (2022, p. 39), un sistema NMT puede traducir correctamente una expresión idiomática, pero no lo hace por una decisión consciente del contexto y sus implicaciones, sino porque el motor infiere, a partir de los textos de entrenamiento, que es la opción más

³ Los «textos de entrenamiento» se refieren a todos los textos que han servido para entrenar al sistema NMT en algún momento, o bien durante la fase de creación («corpus de entrenamiento»), o bien durante la fase de uso, y de los cuales el sistema ha tenido la oportunidad de aprender.

probable. Esto significa que un fallo en el reconocimiento de los elementos implicados podría afectar a toda la oración que contiene la EMV, generando obstáculos de comunicación derivados de la presencia de estos elementos. Por lo tanto, es fundamental aprovechar adecuadamente los avances en TA neuronal, lo cual también es crucial para el progreso de la TA de las EMV.

2. METODOLOGÍA

A fin de analizar el comportamiento y la efectividad que los STA presentan en la actualidad con respecto a las EMV en el par de lenguas español-inglés, fueron necesarios tres componentes esenciales: un corpus monolingüe, herramientas de gestión de corpus para la extracción del objeto de estudio y una selección de traductores automáticos.

2.1. *Corpus*

Para la realización de este estudio, era necesario contar con ejemplos reales de textos turísticos que someter a un proceso de traducción automática, tal y como lo podría haber hecho un usuario habitual de estas herramientas. Con este propósito, se compiló un corpus monolingüe en español de 30 folletos y guías turísticas de distintas regiones españolas que se centrasen en la gastronomía de las regiones o eventos relacionados con esta.

La confección del corpus se llevó a cabo utilizando el protocolo de compilación propuesto por Seghiri (2017, pp. 47-49) y que, posteriormente, se ha implementado en numerosos estudios (Ortego Antón, 2019, 2020, 2024; Fernández Nistal, 2020; Sánchez Carnicer, 2022; Peñuelas Gil, 2024; entre otros). La compilación, por lo tanto, se realizó siguiendo cuatro fases:

- Búsqueda de textos redactados por empresas u organismos que proporcionen servicios turísticos que se centrasen en la vertiente gastronómica de estos y que se ciñesen a la tipología textual a la que se ha acotado el estudio. Para ello, solo se tuvieron en cuenta aquellos textos que aparecían alojados en las páginas web como un documento independiente.
- Descarga, de manera manual, de los textos seleccionados en su formato original (PDF).
- Conversión de los textos a archivos TXT UTF8, para que estos pudiesen ser procesados por herramientas de gestión de corpus, utilizando AntFileConverter (Anthony, 2022).

- Archivo de los textos. En este caso, dado que se trata de un corpus monolingüe, solo se separaron los textos de acuerdo con su formato (PDF o TXT).

A cada texto se le asignó una denominación única basada en el siguiente estándar: identificador numérico según orden de descarga + identificador de la temática (en este caso «GT», como abreviatura de «gastroturismo») + identificador de origen del texto (abreviatura de la provincia o comunidad autónoma para la que se ha escrito) + fecha de descarga (aaaammdd). De este modo, el archivo 001GTSO20230518, por ejemplo, es el primer documento del corpus, y contiene información relevante a la provincia de Soria y se descargó el 18 de mayo de 2023.

El resultado es un corpus monolingüe virtual compuesto por 30 guías y folletos sobre el turismo gastronómico en España –con un número total de palabras (casos o *tokens*) de 93 185 y 16 569 palabras únicas (tipos o *types*)– catalogado y estructurado para su explotación.

2.2. Objeto de estudio

Durante la fase de explotación, se utilizaron dos programas de gestión de corpus –Sketch Engine (Kilgarriff *et al.*, 2004) y AntConc (Anthony, 2020)–, los cuales permitieron analizar y extraer la información pertinente de manera rápida y eficaz. A pesar de que los dos programas presentan herramientas y funcionamientos similares, se optó por utilizar ambas herramientas simplemente por una cuestión de preferencia con respecto a la interfaz de cada una de las herramientas.

El procedimiento de extracción de las EMV se inició utilizando Word List, una de las herramientas básicas de AntConc que permite generar un listado con las palabras únicas del corpus ordenadas según su frecuencia de aparición. Para asegurarnos de que los resultados eran lingüísticamente relevantes, se utilizó una *stoplist* para que el programa ignorase las palabras de menor relevancia como pueden ser artículos, preposiciones o conjunciones. Los resultados se revisaron uno a uno a través de la herramienta N-grams de Sketch Engine, la cual permite filtrar y anidar los resultados, las posibles combinaciones de cada término, prestando siempre atención a la frecuencia.

Este proceso de observación reveló lo común que resulta utilizar expresiones construidas en torno al término «punto», con un total de 44 apariciones en el corpus vinculadas al turismo español. Las EMV detectadas en el corpus que incluyen «punto» presentan características que las hacen interesantes para su análisis:

- Algunas de ellas presentan cierto grado de idiomática, lo que implica que: a) su significado no se puede inferir directamente del significado de sus componentes, sino a través del conocimiento sociocultural de la lengua; y b) es probable que existan diferencias sustanciales con sus equivalentes en lengua inglesa.
- Su composición gramatical posee distintos patrones (sust.+sust., sust.+prep.+sust., sust.+adj, etc.), lo que nos lleva a considerar las posibles diferencias en los resultados.

No obstante, se detectó que cinco de las repeticiones no guardan relación con el turismo gastronómico en ninguna de sus vertientes, sino que tenían un trasfondo histórico o geográfico, por lo que en esta ocasión no se tendrán en cuenta.

Además, otra de las oraciones en las que aparecía «punto» quedó descartada puesto que, aunque hace referencia al punto de cocción, parte de la EMV se omitía por sobreentenderse en la situación («Se prepara a la parrilla y se sirve generalmente poco hecho. Este punto hay a quien le produce cierto rechazo por dar la impresión de venir ensangrentado»).

Al tener en cuenta estas consideraciones, se decidió proceder al análisis utilizando como base las siguientes EMV:

Expresiones multiverbales	Frec.
a punto de nieve	1
aguja de hacer punto	1
en tu punto	1
estar a punto	3
punto caramelo	1
punto de congelación	1
punto de encuentro	1
punto de hebra fina	1
punto de referencia gastronómica	1
punto de sal	7
punto deseado de cocción	1
punto fuerte	15

punto idóneo de cocción	1
punto óptimo de maduración	1
tener su punto	1
un punto picante	1

Tabla 1: Muestra de las EMV del análisis

Fuente. Elaboración propia

2.3. Selección de motores de traducción automática

Durante la selección de los STA que forman parte de este estudio, se tuvieron en cuenta dos puntos básicos: por un lado, el hecho de que, a pesar de que dos motores puedan pertenecer a un mismo paradigma, todos ellos se habrán entrenado con un corpus diferente (Bowker, 2023, p. 101), por lo que los resultados obtenidos serán distintos; y, por el otro, que fuesen representativos de los sistemas que un usuario cualquiera pudiese utilizar, es decir, que sean lo suficientemente conocidos como para que el usuario considere recurrir a ellos. Por supuesto, la popularidad varía dependiendo del par de lenguas, pues no todos los motores traducen a todas lenguas y, en algunos casos, un par de lenguas concreto solo está disponible a través de una lengua intermedia.

Analizando la literatura que trata los STA más populares o hace alusión a ellos (Carré *et al.*, 2022; Rivera-Trigueros, 2022; Bowker, 2023) y contrastando los listados hallados, aislamos cuatro traductores automáticos:

- DeepL (DL), un sistema basado en tecnología NMT;
- Google Translate (GT), también un sistema NMT;
- Microsoft Translator (MST), que utiliza un enfoque híbrido; y
- Yandex.Translate (YT), sistema SMT en el par de lenguas ES-EN⁴.

Los cuatro motores de traducción dan la opción de traducir directamente del español al inglés (estadounidense) sin hacer uso de lenguas pivote. Esto, sumado a su popularidad y variedad, nos hace considerar que conforman una base sólida de cara a comparar los pros y contras de cada uno de ellos. Además, todos ofrecen la posibilidad de trasvasar imágenes en

⁴ El último comunicado de la empresa con respecto a la arquitectura de su STA (Yandex, 2017) indica que se iba a comenzar a utilizar un modelo híbrido, pero únicamente en el par de lenguas inglés-ruso; no hay indicaciones de que esto haya cambiado, por lo que se considerará un traductor automático que funciona con una arquitectura completamente estadística.

su versión aplicación de móvil y archivos completos en su versión web, hechos que podrían potenciar su uso en un contexto turístico.

3. ANÁLISIS Y RESULTADOS

La fase de análisis, por lo tanto, se llevó a cabo utilizando 38 fragmentos, relacionados con el turismo gastronómico, que contienen una EMV compuesta a partir del término «punto», y los cuatro traductores automáticos arriba indicados. Esto implica que en esta fase se han revisado un total de 152 traducciones.

Como se explicaba anteriormente, el propósito del análisis es determinar las variaciones entre las distintas arquitecturas –que un usuario cualquiera podría acabar utilizando– con respecto al uso de EMV, pero también se pretende determinar los posibles obstáculos comunicativos que pueden surgir al emplear estas herramientas. Para ello, se han categorizado los errores detectados siguiendo una modificación del modelo propuesto por Ortiz Boix (2016, pp. 63-64) para la evaluación humana de traducciones, el cual se centra en la precisión (*accuracy*) y la fluidez (*fluidity*) de los resultados obtenidos mediante un proceso de traducción automática.

La modificación consiste en añadir una serie de categorías que se observó eran necesarias para describir los datos del análisis. De este modo, se añadieron dos categorías que afectan directamente a las EMV y dos que, si bien se ven influenciadas por estas, afectan al texto de entrada en general:

- *Mistranslation - grammatical person*: la persona gramatical entre el texto origen (TO) y el texto meta (TM) cambia. Dada la dirección de la traducción (ES>EN), el STA ha desambiguado algunas oraciones de manera incorrecta.
- *Mistranslation - tenses*: el tiempo verbal entre el TO y el TM difiere, lo que cambia el significado del texto.
- *Wrong suggestion*: el STA detecta una estructura que no reconoce en la LO dentro de ese contexto concreto y la sustituye por otra de apariencia similar en la LM, acompañado de la sugerencia «quizás quisiste decir».
- *Repetition*: el STA devuelve una traducción en la que al menos un fragmento del TO se duplica.

Las dos primeras las veremos durante la descripción de los resultados del análisis (véase 3.1.1.), mientras que las dos últimas se comentarán en el resumen de rendimiento de cada STA (véase 3.1.2.).

Por lo tanto, el sistema de categorías utilizado quedó de la siguiente forma:

Accuracy categories	Fluidity categories
<ul style="list-style-type: none"> • Terminology • Mistranslation <ul style="list-style-type: none"> • Overly literal • False friend • Should not have been translated • Date/time • Unit conversion • Number • Entity • Grammatical person • Tenses • Omission • Addition • Untranslated • Wrong suggestion 	<ul style="list-style-type: none"> • Spelling <ul style="list-style-type: none"> • Capitalisation • Diacritics • Typography <ul style="list-style-type: none"> • Punctuation • Unpaired elements • Grammar <ul style="list-style-type: none"> • Morphology • Part of speech • Agreement • Word order • Function words • Unintelligible • Repetition

Tabla 2: Modelo de categorización de errores basado en Ortiz Boix (2016)

Fuente. Elaboración propia

Tras señalar, clasificar y procesar los errores en un programa de hojas de cálculo, donde cada error se contabilizaba individualmente, se procedió a su análisis y comparación. De cara a realizar un análisis completo, se procesaron los datos resultantes en dos fases: en la primera, se examinaron los datos de la traducción de las EMV concretas desde la perspectiva de los STA, tanto de manera individual como en conjunto; y, en la segunda, se analizaron los datos utilizando una abstracción de la estructura gramatical de las EMV.

3.1. Resultados

Al concluir la fase de análisis, se comprobó que 99 de las 152 traducciones revisadas, es decir, el 65,132 % de ellas, no contenían ningún error de traducción en las EMV contempladas y 21 de las 38 oraciones no presentaban errores en ninguno de los motores de TA empleados. Estos datos implican que los 73 errores contabilizados, entre los que se distinguen tanto errores de precisión como de fluidez, se concentran en tan solo 53 de

las traducciones y en 17 de las oraciones. Además, resulta interesante comprobar que solo 28 de las traducciones (9 de DeepL, 6 de Google Translate, 7 de Microsoft Translator y 6 de Yandex.Translate) no albergaron errores de ningún tipo ni en la EMV ni en el contexto inmediato de esta –entendemos por «contexto inmediato» las palabras que rodean a la EMV y que pueden incidir en el significado de esta, como podría ser el caso de «[su punto fuerte]»–.

Cabe mencionar que no se registraron errores en todas las categorías. En lo referido a errores de precisión, fueron siete las categorías sin representación en la muestra de este estudio: *false friend*, *should not have been translated*, *date/time*, *unit conversion*, *number*, *entity* y *grammatical person*. Por otra parte, los errores de fluidez se localizan más en el contexto inmediato a las EMV que en las propias expresiones. De hecho, al analizar exclusivamente las EMV, tan solo se hallaron errores relativos a las categorías *function words*, *part of speech* y *agreement*. Además de estas categorías, también se tratarán *wrong suggestion* y *repetition*, puesto que, si bien su influencia no afecta exclusivamente a las EMV, sí que han tenido un efecto adverso en estas.

A continuación, se presentan los resultados concretos con base en las EMV desde un plano general, específico al STA y a la estructura gramatical.

3.1.1. Errores en las expresiones multiverbales

En lo referido a la traducción de las EMV, la mayoría de los errores se concentran en las categorías de precisión, lo cual se alinea con las expectativas que se tenían, pues la brevedad de las EMV dificulta la aparición de errores de fluidez. A pesar de esto, se registraron cuatro errores de fluidez, los cuales se dividen en tres categorías distintas: uno en *part of speech*, uno en *agreement* y dos en *function words*.

Estos dos últimos se dan en la oración «Damos punto de sal, colamos y reservamos», donde el verbo utilizado por dos de los STA requiere una referencia al objeto directo para la correcta estructuración de la frase (véase la Tabla 3). Las traducciones que contienen los errores han sido proporcionadas por los TA de Microsoft y Yandex, los dos sistemas seleccionados que trabajan principalmente con información estadística, por lo que el error podría estar motivado por cuestiones del método de procesamiento que utilizan. Además, existen dos factores que podrían contribuir a esto: por un lado, el hecho de que la frase original no presente artículos y, por el otro, la tendencia más elevada a la traducción palabra por palabra que se ha observado en sus traducciones.

DL	Add salt , strain and reserve.
GT	Season with salt , strain and reserve.
MT	Give point of salt , strain and reserve.
YX	We give point of salt , strain and reserve.

Tabla 3: Comparación traducción «damos punto de sal»

Fuente. Elaboración propia

Por su parte, los sistemas que trabajan exclusivamente con tecnología neuronal han encontrado soluciones alternativas al problema, haciendo uso de dos estructuras y expresiones comunes en un registro informal del inglés y típicas en el ámbito gastronómico.

No obstante, los otros dos errores de fluidez se dan en el sistema DeepL en relación con la EMV «un punto picante», donde la oración «[...] con fritada, un punto picante o con huevos» se ha traducido como *with fried, spicy or with eggs*, conformando además el único caso de omisión parcial de la EMV detectado en la muestra. En esta ocasión, se contemplaba la posibilidad de que fuesen factores externos a la expresión los que interferían con el proceso de traducción, pues DeepL no había tenido problemas en traducir expresiones de características similares en otros contextos, y la EMV se da en una enumeración dentro de otra enumeración y en una oración con una gran densidad terminológica. Estos factores, unidos a la tendencia de todos los sistemas a ser demasiado literal, podrían haber desembocado en los errores perceptibles en este ejemplo.

Cabe mencionar que la tendencia a ser demasiado literal, si bien está presente en todos los sistemas, se ha dado en distintas medidas. Este tipo de error se ha detectado en 4 ocasiones en DeepL, 7 en Google Translate, 15 en Microsoft Translator y 13 en Yandex.Translate, lo que convierte a *overly literal* en la categoría más común en lo referido (estrictamente) a la traducción de las EMV con un total de 39 casos. La traducción demasiado literal supone el 53,425 % del total de los errores (véase el Gráfico 1), lo que resalta la clara tendencia a traducir los elementos que componen una expresión de manera aislada. Sin embargo, este método resulta, en general, incompatible con las características intrínsecas de las EMV.

Para ponerlo en perspectiva, el siguiente error más común dentro de la muestra de estudio tan solo cuenta con 13 ocurrencias (17,808 %) y se trata de errores de terminología.

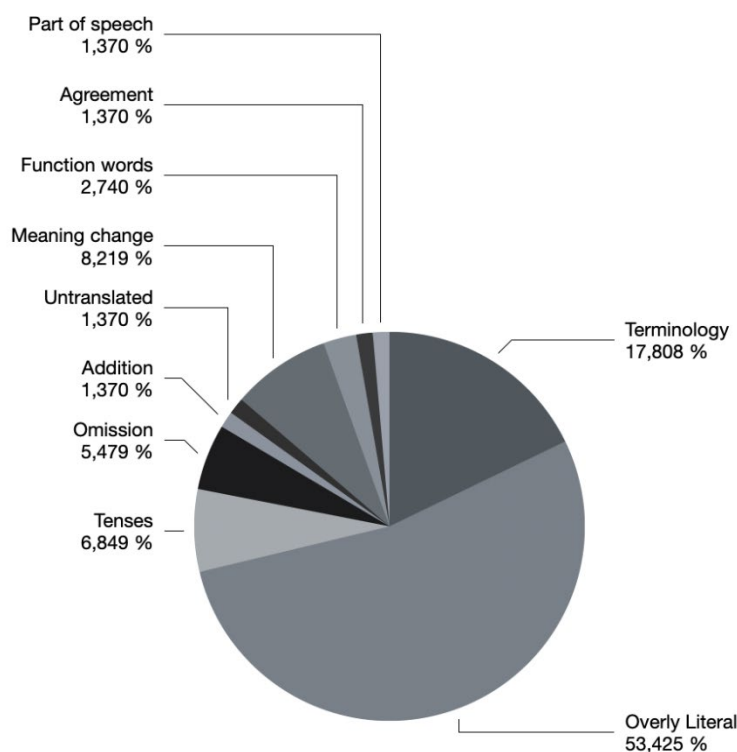


Gráfico 1. Representación en porcentajes de los errores hallados en la traducción de EMV

Fuente. Elaboración propia

Conviene destacar el hecho de que los errores de terminología, así como los de tiempos verbales, que cambian el significado de la frase, en muchos casos vienen dados por esa tendencia a la literalidad que mencionábamos anteriormente. De hecho, se ha comprobado que, en el 84,615 % de los casos analizados, los errores de terminología están relacionados con la traducción demasiado literal. Sobre todo, ocurre en el caso de la EMV «[a] punto [de] caramelo», donde los fallos de redacción en la LO suponen una dificultad añadida. Por lo tanto, no es de extrañar que, además de la tendencia a la traducción palabra por palabra en esta situación, los STA hayan optado por trasvasar «caramelo» como *candy* (Yandex) o *stitch* (DeepL), ambas traducciones correctas, si bien en otros ámbitos.

Sucede algo similar con los tiempos verbales: el no comprender las expresiones como elementos idiomáticos ha afectado la forma en la que los

sistemas procesan las EMV y el contexto inmediato de estas, lo que modifica su significado. Los tres casos contemplados en la categoría *tenses* se han dado en la EMV «estar a punto» (véase la Tabla 4). En dos de ellos, el verbo introductorio ha quedado completamente opacado por la EMV, traduciendo así «cuando vaya a estar a punto» por *when it is ready* (DeepL y Google Translate), que presenta una realidad similar, pero distinta y que se aleja de la precisión propia del lenguaje culinario. La tercera se da cuando la expresión cobra el sentido de ‘está a punto de llegar’, el cual se infiere al expandir el contexto a la oración siguiente (en corchetes en la Tabla 4), pero que Microsoft Translator ha traducido por *is ready* indistintamente de si se añadía un contexto más amplio o no.

Original	Cuando vaya a estar a punto, se retira del fuego.
DL	When it is ready , it is removed from the heat.
GT	When it is ready , it is removed from the heat.
Original	Lo mejor está a punto. [Es la época de pescarlas con la medida de rigor... son las solicitadas antxoas del Cantábrico, que llegan con los saludables verdes.]
MT	The best is ready . [It is the time to fish them with the measure of rigor... are the requested antxoas of the Cantabrian, which arrive with the healthy verdes.]

Tabla 4: Errores de tiempo verbal

Fuente. Elaboración propia

También observamos que Google Translate es el único sistema que presenta un error de adición. En esta ocasión, el STA ha matizado que es necesario añadir *a bit of salt* cuando, en la frase original («... se les añade el punto de sal...»), no se menciona la cantidad. Lo más probable es que el error provenga de haber interpretado «punto» como ‘unidad de medida’; aunque este se trata de un significado poco común.

El caso opuesto, la omisión de contenido, se puede observar en cuatro ocasiones. Concretamente, en la muestra de estudio se ha hallado un único caso de omisión parcial, el cual se ha mencionado con anterioridad al tratar la expresión «un punto picante». En esta ocasión, DeepL, en vez de traducir la EMV en su totalidad, simplemente hizo referencia al «picante» y sin tener en cuenta la estructura gramatical de la frase a la que se añadía. Los otros tres ejemplos identificados corresponden a omisiones completas de la expresión y se han producido en dos EMV distintas. La primera es «a punto

de nieve [muy firme]». En este caso, Yandex optaba por hablar de la consistencia final deseada, arriba indicada entre corchetes, sin mencionar en ningún momento el equivalente de la EMV (*stiff peaks*). La segunda expresión es «en tu punto», donde tanto DeepL y Google Translate han decidido obviar la referencia al gusto del consumidor que existe en el original. Esta expresión, también ha supuesto un obstáculo para los otros STA, donde Microsoft Translator ha optado por la traducción palabra por palabra y Yandex.Translate la ha dejado sin traducir.

Una vez presentados los resultados generales de la traducción automática de EMV, procedemos a analizar brevemente lo que esto implica desde el punto de vista de cada motor de traducción.

3.1.2. Desglose del rendimiento por sistema de traducción automática

Como muestra la Tabla 5, no existen grandes diferencias en el rendimiento de cada uno de los traductores automáticos dado el tamaño de la muestra de análisis y el enfoque del estudio. A pesar de ello, se han identificado cuestiones significativas en lo que respecta al número y tipo de los errores. Al procesar los errores de cada STA, se ha observado que Google Translate es el que obtiene mejores resultados, si bien el margen es mínimo al compararlo con el resto de los sistemas:

	Errores
DeepL	17
Google Translate	16
Microsoft Translator	19
Yandex.Translate	21

Tabla 5: Resumen de errores por STA

Fuente. Elaboración propia

El hecho de que las cifras sean tan similares entre los cuatro traductores se debe a que, si bien los dos sistemas neuronales cometen de media menos errores por categoría, registran errores en un mayor número de categorías (véase la Tabla 6). Por el contrario, las cifras que presentan Microsoft Translator y Yandex.Translate son más dispares, siendo particularmente notorio el caso de *overly literal* (donde estos acumulan 15 y 12 errores, respectivamente). No obstante, el resto de las categorías presentan errores mínimos o, en el caso de Microsoft, registran fallos en un menor número de categorías.

Centrándonos en los resultados de DeepL, observamos que se trata del sistema que menos errores por traducción literal presenta. Sin embargo, a pesar de sus buenos resultados en esta categoría, es el sistema en el que se contabilizan más fallos de terminología (5) y de omisión (2). Asimismo, es el único en el que se han hallado dos casos de repetición, entendida como la duplicación de un fragmento del TO. Estos errores no se han dado en referencia a una EMV relacionada con «punto», por lo que no se han tenido en cuenta para el recuento final de este estudio. Sin embargo, sí están relacionados con fragmentos en los que hay un número significativo de expresiones multiverbales, lo cual es algo a considerar. Por ejemplo:

[...] calf's tongue in red wine, liver with onions, pig's hands, pig's ears, which, like those of the lambs, are eaten in batter as a snack in a stew, veal tongue in red wine, liver with onions, pig's trotters, pig's ears, which, like those of the lambs, are eaten in batter as a snack in bars [...].

Otro error exclusivo de un sistema de traducción automática, esta vez Google Translate, es la sugerencia de modificaciones incorrectas en el TO con la forma «Quizás quisiste decir: [...]». Se han detectado 5 fallos en total de estas características, los cuales se estudiaron con relación al contexto, pues afectaban a la frase en su conjunto más que a la EMV de manera concreta. De los 5 errores, uno se refería a un cambio de la LO al catalán, que ofrecía resultados de traducción similares, pero no idénticos, e incluía errores de precisión; uno, a la persona gramatical; uno, a un cambio terminológico («vocación micoturística» se sustituye por «vocación ecoturística»); y tres, al verbo introductorio de la EMV, lo que generó problemas de terminología y de palabras estructurales (p. ej.: «los cocemos en el caldo hasta que estén a punto» por «nos conocemos en el caldo hasta que estén a punto»).

En contraste, ni Yandex.Translate ni Microsoft Translator presentan errores específicos de sus sistemas por los que haya sido necesario crear nuevas categorías. Si bien su tendencia a la literalidad es muy superior a la de los otros dos traductores automáticos, especialmente si se tiene en cuenta también el contexto inmediato de las EMV. Esto nos lleva a considerar las discrepancias existentes entre el uso de estos dos motores y las características de las expresiones multiverbales. No obstante, cabe destacar los buenos resultados obtenidos por Microsoft Translate en el resto de las categorías, pues se trata del motor que menos tipos de error comete, al registrar errores en solo tres de ellas.

	DL	GT	MT	YX
Terminology	5	2	3	3
Overly literal	4	7	15	13
Tenses	2	2	0	1
Meaning change	2	3	0	1
Omission	2	1	0	1
Addition	0	1	0	0
Untranslated	0	0	0	1
Part of speech	1	0	0	0
Agreement	1	0	0	0
Function words	0	0	1	1

Tabla 6: Desglose de errores en la traducción de EMV por STA

Fuente. Elaboración propia

3.1.3. Errores según su categoría gramatical

Un rasgo interesante para analizar es si la composición gramatical de cada EMV puede afectar al número de errores que se cometen en su traducción. Para ello, se identificó la estructura de cada una de las 16 expresiones que forman parte de este trabajo (véase la Tabla 1) y se cuantificó el número total de ocurrencias de cada estructura –es decir, se tomó en cuenta la frecuencia de aparición– y el número de errores que se habían detectado en cada caso.

Los datos que se proporcionan en la Tabla 7 muestran que dos de las estructuras gramaticales no presentan errores de traducción. Esto resulta muy significativo al considerar que una de estas estructuras, *sustantivo + adjetivo* («punto fuerte»), es la más repetida en el corpus con un total de 15 ocurrencias.

	EMV únicas	Total de casos	Errores
art. + sust. + adj.	1	1	5
prep. + pron. + sust.	1	1	4
prep. + sust. + prep. + sust.	1	2	2
sust. + adj.	1	15	0

sust. + adj. + prep. + sust.	3	3	14
sust. + prep. + sust.	3	10	22
sust. + prep. + sust. + adj.	2	2	6
sust. + prep. + verb. + sust.	1	1	0
sust. + sust.	1	1	6
verb. + prep. + sust.	1	3	10
verb. + pron. + sust.	1	1	4

Tabla 7: Desglose de errores por estructura gramatical

Fuente. Elaboración propia

En el caso opuesto, tenemos estructuras gramaticales que, a pesar de aparecer solo una vez en la muestra de estudio, contienen errores en 3 o 4 de los motores, como es el caso de:

- artículo + sustantivo + adjetivo («un punto picante»)
- preposición + pronombre + sustantivo («en tu punto»)
- sustantivo + sustantivo («punto caramelo»)
- verbo + preposición + sustantivo («estar a punto»)
- verbo + pronombre + sustantivo («tener su punto»)

Otro caso que destacar es el de *sustantivo + preposición + sustantivo*, que, con nueve ocurrencias, constituye una de las combinaciones más comunes y en la que se ha hallado un total de 22 errores. Esto implica que, de media, 2,2 de los STA fallan en cada una de las 10 oraciones que contienen una EMV con esta estructura, un número elevado si se contrasta con el resto de los casos estudiados. Una situación similar se observa en el caso de *sustantivo + adjetivo + preposición + sustantivo*, pues, aunque cuenta con solo tres ocurrencias, presenta siete errores.

No obstante, el número de errores no es lo suficientemente elevado como para hacer afirmaciones categóricas al respecto, pero sí que parece indicar que existen unos patrones de comportamiento en función de la estructura gramatical de la expresión que debe ser traducida en los que sería interesante profundizar en trabajos futuros.

CONCLUSIONES

Los traductores automáticos se han convertido en una herramienta extremadamente popular en muchos de los ámbitos de la sociedad, entre

ellos el turismo en todas sus variantes. Sin embargo, a pesar de los avances, los sistemas de traducción automática aún presentan grandes carencias que no siempre se tienen en cuenta al utilizarlos, lo que lleva a textos cuya aceptabilidad se podría poner en duda y lo que supone un riesgo añadido para el sector.

El fin último de este trabajo era analizar el comportamiento y la efectividad de la traducción automática con respecto a un objeto de estudio concreto, las expresiones multiverbales, puesto que se trata de elementos lingüísticos extremadamente comunes en el discurso turístico y que, dadas sus implicaciones idiomáticas, pueden afectar la comunicación seriamente. Esta circunstancia se ve potenciada al tener en cuenta el funcionamiento de los STA, los cuales se basan en la tecnología de corpus para calcular la traducción más adecuada. En consecuencia, un cambio en un elemento fijo como son las EMV puede suponer modificaciones significativas en el contexto que las rodea.

Los resultados del estudio han mostrado una clara tendencia por parte de los sistemas analizados a la traducción literal o palabra por palabra en lo referido a las EMV, particularmente en el caso de los sistemas que trabajan con tecnología estadística. Además, también se ha encontrado una propensión a errores que afectan principalmente a la integridad estructural de los textos (errores de persona gramatical, tiempos verbales, etc.) y al contenido en sí mismo (terminología errónea y términos que no se deberían traducir y viceversa, omisiones, adiciones, etc.).

Es previsible que los textos turísticos aporten información sobre una cultura que difiere, en mayor o menor medida, de la vida cotidiana del turista, por lo que se espera que los textos no solo estén traducidos a su lengua, sino que sean precisos, fáciles de comprender y que respondan a las expectativas derivadas de factores culturales que el turista pueda tener. Se trata de cuestiones que se ven opacadas por los errores aquí abordados y, que, a la larga, podrían resultar contraproducentes desde el punto de vista económico, dado que, como se mencionaba en la introducción, el nivel de exigencia por parte del turista aumenta paulatinamente. Sin embargo, en este trabajo se han empezado a atisbar ciertos patrones de procesamiento en lo que a la traducción de EMV respecta, así como potenciales cuestiones de estructura gramatical que podrían motivar algunos de estos cambios.

De cara al futuro, sería interesante ampliar el número y la variedad de EMV analizadas para poder contar con una base de datos que permita realizar afirmaciones concluyentes, puesto que las que aquí se recogen son solo una pequeña muestra de la realidad fraseológica en el turismo. Además, sería necesario ampliar esta metodología a otros elementos lingüísticos

característicos del ámbito turístico. Esto permitiría describir con mayor precisión el rendimiento de estas herramientas, identificar sus carencias y puntos fuertes, y proponer soluciones a los problemas que plantean.

REFERENCIAS BIBLIOGRÁFICAS

- Álvarez Jurado, M. (2020). Adquisición y transmisión del conocimiento experto a través de la traducción de las guías turísticas de arquitectura. *Onomázein*, (NE VII), 1-17. <https://doi.org/10.7764/onomazein.ne7.01>
- Anthony, L. (2020). AntConc (Versión 3.5.9) [Programa de ordenador]. Waseda University. <https://www.laurenceanthony.net/software/antconc/>
- Anthony, L. (2022). AntFileConverter (Version 2.0.2) [Programa de ordenador]. Waseda University. <https://www.laurenceanthony.net/software>
- Austermühl, F. y Kortenbuck, A. (2012). A translator's sword of Damocles? An introduction to machine translation. En F. Austermühl, *Electronic tools for translation* (3ª ed, pp. 153-176). Routledge.
- Biber, D., Johansson, S., Leech, G., Conrad, S. y Finegan, E. (1999). *Grammar of spoken and written English*. Pearson Education Limited. <https://doi.org/10.1075/z.232>
- Bowker, L. (2023). *De-mystifying translation: introducing translation to non-translators*. Routledge. <https://doi.org/10.4324/9781003217718>
- Bowker, L. y Buitrago Ciro, J. (2019). *Machine translation and global research: towards improved machine translation literacy in the scholarly community*. Emerald Publishing.
- Brown, P., Cocke, J., Della Pietra, S., Della Pietra, V., Jelinek, F., Lafferty, J. D., Mercer, R. y Roossin, P. (1990). A statistical approach to language translation. *Computational linguistics*, 16(2), 79-85. <https://aclanthology.org/J90-2002>
- Brown, P., Cocke, J., Della Pietra, S., Della Pietra, V., Jelinek, F., Mercer, R. y Roossin, P. (1988). A statistical approach to language translation. *Proceedings of the 12th conference on Computational linguistics*, 1, pp. 71-76. <https://aclanthology.org/C88-1016>
- Carré, A., Kenny, D., Rossi, C., Sánchez-Gijón, P. y Torres-Hostench, O. (2022). Machine translation for language learners. En D. Kenny (Ed.), *Machine translation for everyone: empowering users in the age of*

- artificial intelligence (pp. 187-207). Language Science Press. <https://doi.org/10.5281/zenodo.6760024>
- Corpas Pastor, G. (2013). Detección, descripción y contraste de las unidades fraseológicas mediante tecnologías lingüísticas. En I. Olza Moreno y E. Manero Richard (Coords.), *Fraseopragmática* (pp. 335-374). Frank & Time.
- Costa-Jussà, M. R. y Fonollosa, J. A. (2015). Latest trends in hybrid machine translation and its applications. *Computer Speech and Language*, 32(1), 3-10. <https://doi.org/10.1016/j.csl.2014.11.001>
- Fernández Nistal, P. (2020). Los corpus como herramienta de traducción para los traductores del siglo XXI: el caso del chorizo ibérico de bellota. En S. Álvarez Álvarez y M. T. Ortego Antón (Eds.), *Perfiles estratégicos de traductores e intérpretes. La transmisión de la información experta multilingüe en la sociedad del conocimiento del siglo XXI* (pp. 143-160). Comares.
- Jackendoff, R. (1997). *The architecture of the language faculty*. MIT Press. <https://doi.org/10.2307/417010>
- Kenny, D. (2022). Human and machine translation. En D. Kenny (Ed.), *Machine translation for everyone: empowering users in the age of artificial intelligence* (pp. 23-49). Language Science Press. <https://doi.org/10.5281/zenodo.6653406>
- Kilgarrieff, A., Rychlý, P., Smrž, P. y Tugwell, D. (2004). The Sketch Engine. *Proceedings of the 11th EURALEX International Congress*, pp. 105-116.
- Koehn, P., Och, F. J. y Marcu, D. (2003). Statistical phrase-based translation. *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 127-133. <https://aclanthology.org/N03-1017>
- Ministerio de Industria, Comercio y Turismo. (2023, 2 de febrero). Datos de Frontur y Egatur del INE [Nota de prensa]. <http://www.mincotur.gob.es/es-es/GabinetePrensa/NotasPrensa/2023/Paginas/En-2022-visitaron-España-71,6-millones-de-turistas-internacionales-que-realizaron-un-gasto-de-87.061-millones-de-euros.aspx>
- Mitkov, R., Seretan, V., Corpas Pastor, G. y Monti, J. (2018). Multiword units in machine translation and translation technology. En R. Mitkov, V. Seretan, G. Corpas Pastor y J. Monti (Eds.), *Multiword units in machine*

- translation and translation technology (pp. 1-38). John Benjamins. <https://doi.org/10.1075/cilt.341.01mon>
- Moorkens, J. (2022). Ethics and machine translation. En D. Kenny (Ed.), *Machine translation for everyone: empowering users in the age of artificial intelligence* (pp. 121-140). Language Science Press.
- Oliver, A. (2016). *Herramientas tecnológicas para traductores*. Editorial UOC.
- Organización Mundial del Turismo. (2023). *World tourism barometer - May 2023 (excerpt)*, 21(2). https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2023-05/UNWTO_Barom23_02_May_EXCERPT_final.pdf
- Ortego Antón, M. T. (2019). La terminología del sector agroalimentario (español-inglés) en los estudios contrastivos y de traducción especializada basados en corpus: los embutidos. Peter Lang. <http://doi.org/10.3726/b15808>
- Ortego Antón, M. T. (2020). Las fichas descriptivas de embutidos en español y en inglés: un análisis contrastivo de la estructura retórica basado en corpus. *Revista Signos*, 53(102), 170-194. <http://doi.org/10.4067/S0718-09342020000100170>
- Ortego Antón, M. T. (2024). The design of Torrezno TRAD: the semiautomatic Spanish-English writing and translation aid tool. En I. Peñuelas Gil y M. T. Ortego Antón (Eds.), *Interpreting and translation for agri-food professionals in the global marketplace* (pp. 69-84). De Gruyter. <https://doi.org/10.1515/9783111101729-004>
- Ortiz Boix, C. (2016). *Implementing machine translation and post-editing to the translation of wildlife documentaries through voice-over and off-screen dubbing* [Tesis doctoral, Universitat Autònoma de Barcelona]. <http://hdl.handle.net/10803/400020>
- Penadés Martínez, I. (2015). *Para un diccionario de locuciones. De la lingüística teórica a la fraseografía práctica*. Universidad de Alcalá.
- Peñuelas Gil, I. (2024). *Estudio contrastivo del tratamiento de las expresiones multiverbales del turismo gastronómico en los sistemas de traducción automática del español al inglés* [Tesis doctoral, Universidad de Valladolid]. <https://doi.org/10.35376/10324/67810>
- Pérez Blanco, M. e Izquierdo, M. (2021). Developing a corpus-informed tool for Spanish professionals writing specialized texts in English. En J. Lavid-López, C. Maíz-Arévalo y J. R. Zamorano-Mansilla (Eds.),

- Corpora in translation and contrastive research in the digital age (pp. 147-173). John Benjamins. <https://doi.org/10.1075/btl.158.06per>
- Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: a systematic review. *Lang Resources & Evaluation*, 56, pp. 593–619. <https://doi.org/10.1007/s10579-021-09537-5>
- Sag, I. A., Baldwin, T., Bond, F., Copestake, A. y Flickinger, D. (2002). Multiword expressions: a pain in the neck for NLP. Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002), pp. 1–15. https://doi.org/10.1007/3-540-45715-1_1
- Sánchez Carnicer, J. (2022). Traducción y discapacidad. Un estudio comparado de la terminología inglés-español en la prensa escrita. Peter Lang. <https://doi.org/10.3726/b19567>
- Sánchez Ramos, M. M. y Rico Pérez, C. (2020). Traducción automática: conceptos clave, procesos de evaluación y técnicas de posesición. Comares.
- Seghiri, M. (2017). Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores. *Babel*, 63(1), 43-64. <https://doi.org/10.1075/babel.63.1.04seg>
- Vieira, L. N. (2020). Machine translation in the news: a framing analysis of the written press. *Translation Spaces*, 9(1), pp. 98–122. <https://doi.org/10.1075/ts.00023.nun>
- Yamada, K. y Knight, K. (2001). A syntax-based statistical translation model. Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, pp. 523-530. <https://aclanthology.org/P01-1067>
- Yandex. (2017, 14 de septiembre). One model is better than two. Yandex.Translate launches a hybrid machine translation system. <https://yandex.com/company/blog/one-model-is-better-than-two-yu-yandex-translate-launches-a-hybrid-machine-translation-system>
- Zens, R., Och, F. J. y Ney, H. (2002). Phrase-based statistical machine translation. En M. Jarke, J. Koehler y G. Lakemeyer (Eds.), *KI 2002. Advances in Artificial Intelligence: 25th Annual German Conference on AI, KI 2002 Aachen, Germany, September 16–20, 2002 Proceedings*, 2479 (pp. 18–32). Springer. https://doi.org/10.1007/3-540-45751-8_2